

Fig 3

Amino acids said to be 'similar' are: A,S,T; D,E; N,Q; R,K; I,L,M,V; F,Y,W

AAV2VP1 - MAADGYLPDWLEDTLSEGIQWWKLKPGPPPKPAERHKDDSRGLVLPGYKYLGP -55
 AAV4VP1 - MT-DGYLPDWLEDNLSEGVREWWALQPGAPKPKANQQHQDNARGLVLPGYKYLGP -54
 AAV2VP1 - FNGLDKGEPVNEADAAALEHDKAYDRQLDSGDNPYLKYNHADADEFQERLKEDTSF -110
 AAV4VP1 - GNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQQRLOGDTSF -109
 AAV2VP1 - GGNLGRAVFQAKKRVLEPLGLVEEPVKTAPGKKRPVEHSPVEPDSSSGTGKAGQQ -165
 AAV4VP1 - GGNLGRAVFQAKKRVLEPLGLVEQAGETAPGKKRPLIESPQQPDSSTGIGKKKGKQ -164
 AAV2VP1 - PARKRLNFGQ-TGDADSVDPDQPLGQPPAAPSGGLGTNTMATGSGAPMADNNE--- -216
 AAV4VP1 - PAKKKLVFEDETGAGDGPPEG-----STSGAMSDDSEMRA -199
 AAV2VP1 - -----GADGVGNSSGNWHCDSTWMGDRVITTTSTRTWALPTYNNHLYKQI -260
 AAV4VP1 - AAGGAAVEXGQADGVGNASGDWHCDSTWSEGHVTTTSTRTWVLPPTYNNHXYKRL -254
 AAV2VP1 - SSQSGASNDNH^{VQ}YFGYSTPWGYFDNRFHCHFSPRDWQRLINNNWGFPRKRLNFKL -315
 AAV4VP1 - ---GESLQSN^{VQ}TYNGFSTPWGYFDNRFHCHFSPRDWQRLINNNWGM^HPRKAMRVKI -306
 AAV2VP1 - FNIQVKEVTQNDGTTTIAN^{VQ}NLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADV -370
 AAV4VP1 - FNIQVKEVTTSNGETTVANNLTSTVQIFADSSYELPYVMDAGQEGSLPPFPNDV -361
 AAV2VP1 - MVPQYGY---LTLNNGSOAVGRSSFYCLEYFPSQMLRTGNNFTFSYTFEDVPFHS -422
 AAV4VP1 - MVPQYGYCGLVTGNTSQQQTDRNAFYCLEYFPSQMLRTGNNFEITYSFEKVPFHS -416
 AAV2VP1 - SYAHSQSLDRLMNPLIDQYLYLSRTNTPS----GTTTOSRLQFSQAGASDIRDQ -473
 AAV4VP1 - MYAHSQSLDRLMNPLIDQYLWGLQSTTTGTTLNAGTATTN---FTKL^{VQ}RPTNFSNF -468
 AAV2VP1 - SRNWLPGPCYRQQRVSKTSADNNNSEYSWTGA-----TKYHLNGRDSL^{VQ}VNPGPAM -523
 AAV4VP1 - KKNWLPGPSIKQQGF^{VQ}SKTANQNYKIPATGSDSLIKYETHSTLDGRWSALTPGPPM -523
 AAV2VP1 - ASHKDDEEKFFPQSGVLIFGKQGSEKT^{VQ}NVDIEKVM----ITDEEEI^{VQ}RTTNPVATE -574
 AAV4VP1 - ATAGPADSKFS--NSQLIF---AGPKQNGNTATVPGT^{VQ}LIFTSEEELAA^{VQ}TNATD^{VQ}TD -573
 AAV2VP1 - QYGSV-----STNLQ^{VQ}RGNRQAATADVNTQGVLPGMVWQDRDVYLQGP^{VQ}IAKIPH -623
 AAV4VP1 - MWGNLPGGDQSN^{VQ}SNLPTVDR-----LTALGAVPGMVWQNRDIYYQGP^{VQ}IAKIPH -622
 AAV2VP1 - TDGHFHPSPLMGGFGLKHPPPQILIKNTVPANPSTTFSAAKFASFITQYSTGQV -678
 AAV4VP1 - TDGHFHPSPLIGGFGLKHPPPQIFIKNTVPANPATTFSSTPVNSFITQYSTGQV -677
 AAV2VP1 - SVEIEWELQKENS^{VQ}KRWNPEIQYTSN^{VQ}YNKSVNVDFTVD^{VQ}TNGVYSEPRPIGTRYLTR -733
 AAV4VP1 - SVQIDWEIQKERSK^{VQ}RWNPEVQFTS^{VQ}NYGQQNSLLWAPDAAGKYTEPRAIGTRYLTH -732
 AAV2VP1 - NL -735
 AAV4VP1 - HL -734

Identity : 453 (61.7%)
 Similarity: 64 (8.7%)

— supposedly exposed regions.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows (A) the sequence of AAV4 contained in the BssHIII fragment was derived using an automated sequencer, and (B) the organization of AAV4, showing a map of the genomic organization of AAV4 and the organization of the overlapping clones.

Fig. 2 shows a comparison of Rep ORFs. The Rep open reading frames of AAV2 and AAV4 were compared using the nalign program of Pcgene. Identical amino acids are indicated by a : and similar amino acids are indicated by a single dot.

10

Fig. 3 shows a comparison of Cap ORFs. The Capsid open reading frames of AAV2 and AAV4 were compared using the paln program of Pcgene. Identical amino acids are indicated by a : and similar amino acids are indicated by a single dot. The theoretical initiator codons of VP2 and VP3 are indicated. Large blocks of dissimilarity are highlighted in yellow. Regions which have been proposed to be on the surface of AAV are underlined.

15

Fig. 4 shows the AAV4 p5 promoter. The p5 promoter was identified by comparison to the AAV2 seq. The TATA box and Rep binding sites and initiation codon are identified by boxes. The p5 transcription start site is indicated by an arrow as well as the TRS. The YY1 binding sites are underlined. The unique AAV4 sequence is highlighted nt 209-269).

20

Fig. 5 shows AAV4 ITR. The sequence of the "flip" ITR is shown in the hairpin conformation. The Rep binding site is identified by overlining. The cleavage site in TRS is indicated by an arrow. Bases shown to be important based on methylation interference binding are indicated by dots and asterisks.

25